WHAT IS CLAIMED IS:

1. A packet voice conferencing method comprising:

receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

digitally encoding a signal block to represent the first and second sound field signals as captured during a first time period;

estimating the relative temporal delay between the first and second sound field signals within the approximate timeframe of the first time period;

transmitting to a remote conferencing point, in packet format, both the encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay.

2. The method of claim 1, wherein digitally encoding a signal block comprises combining the first and second sound field signals into a composite sound field signal by a method selected from the group of methods consisting of:

selecting one sound field signal as the source of the composite sound field signal and discarding the other sound field signal;

summing the first and second sound field signals; and

averaging the first and second sound field signals.

3. The method of claim 1, wherein estimating the relative temporal delay comprises:

calculating, for each of a plurality of relative time shifts, a first-to-second sound field signal cross-correlation coefficient; and

selecting the relative temporal delay to correspond to the relative time shift generating

the largest cross-correlation coefficient.

4. The method of claim 3, wherein estimating the relative temporal delay further comprises tracking the beginning and ending of a talkspurt represented in the sound field signals, and limiting the variation of the estimated relative temporal delay during a talkspurt.

5. The method of claim 1, wherein the relative temporal delay associated with the first time period is estimated using substantially only the sound field signals captured during the first time period.

6. The method of claim 1, wherein estimating the relative temporal delay further comprises tracking the beginning and ending of a talkspurt represented in the sound field signals, wherein relative temporal delay associated with the first time period is estimated using substantially all of the sound field signals corresponding to the current talkspurt, up to and including at least a first portion of the first time period.

7. The method of claim 1, wherein estimating the relative temporal delay comprises detecting the beginning time of a talkspurt in each of the sound field signals, and selecting the relative temporal delay for a talkspurt to correspond to the difference in beginning times detected for that talkspurt.

8. The method of claim 1, wherein the stereo decoding parameter expresses the estimated relative temporal delay between the first and second sound field signals as an integer number of digital sampling intervals.

9. The method of claim 1, wherein the stereo decoding parameter expresses an estimated angle of arrival based on the estimated relative temporal delay and the relative positioning of the first and second spatially-separated points.

10. The method of claim 1, wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in the same packet as that sample block.

11. The method of claim 1, wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in a later packet than that sample block.

12. The method of claim 1, wherein the stereo decoding parameter corresponding to the digitally-encoded signal block representing the first time period is transmitted in a packet separate from any digitally-encoded sample block.

13. The method of claim 1, wherein the stereo decoding parameter is transmitted once per talkspurt.

14. The method of claim 1, further comprising estimating the signal energy present in each sound field signal during the approximate timeframe of the first time period, and transmitting to the remote conferencing endpoint, in packet format, a stereo balance parameter related to the relative signal energy in each sound field signal.

15. The method of claim 1, further comprising estimating the signal energy present in a frequency subband of each sound field signal during the approximate timeframe of the first time period, and transmitting to the remote conferencing endpoint, in packet format, a stereo balance parameter related to the relative signal energy in that subband for each sound field signal.

16. The method of claim 1, further comprising establishing a packet-based control protocol with the remote conferencing point, and using the control protocol to inform the remote conferencing point that an encoder performing the method of claim 1 is available for stereo packet voice conferencing.

17. An apparatus comprising a computer-readable medium containing computer instructions that, when executed, cause a processor or multiple communicating processors to perform a method for packet voice conferencing, the method comprising:

receiving concurrently-captured first and second voice sample streams, the first stream representing a first sound field signal captured at a first spatial location within a sound field, the second stream representing a second sound field signal captured at a second spatial location within the sound field;

encoding a block of combined voice samples for the first and second voice sample streams, the block representing voice samples captured during a first time period;

estimating, using voice samples captured in the approximate timeframe of the first time period, the relative temporal delay between the first and second sound field signals;

transmitting to a remote conferencing point, in packet format, both the encoded block

of combined voice samples and a stereo decoding parameter based on the estimated relative temporal delay.

18. The apparatus of claim 17, wherein encoding a block of combined voice samples comprises combining voice samples for the first and second voice sample streams by a method selected from the group of methods consisting of:

selecting one sample stream as the source of combined voice samples and discarding the other;

summing a sample from the first stream and a sample from the second stream, the samples representing substantially the same sample period; and

averaging a sample from the first stream and a sample from the second stream, the samples representing substantially the same sample period.

19. The apparatus of claim 17, wherein estimating the relative temporal delay comprises:

calculating, for each of a plurality of sample index shift distances, a cross-correlation coefficient for a group of samples from one sample stream and a corresponding group of index-shifted samples from the other sample stream; and

selecting the relative temporal delay to correspond to the sample index shift distance generating the largest cross-correlation coefficient.

20. The apparatus of claim 19, wherein estimating the relative temporal delay further comprises tracking the beginning and ending of a talkspurt on the voice sample streams, and limiting the variation of the estimated relative temporal delay during a talkspurt.

21. The apparatus of claim 17, wherein the group of samples from one sample stream comprise the samples captured during the first time period.

22. The apparatus of claim 17, wherein estimating the relative temporal delay further comprises tracking the beginning and ending of a talkspurt on the voice sample streams, wherein the group of samples from one sample stream comprise approximately all samples received within a current talkspurt, up to and including at least a first portion of the first time period, for that sample stream.

23. The apparatus of claim 17, wherein estimating the relative temporal delay comprises detecting the beginning time of a talkspurt in each of the first and second sample streams, and selecting the relative temporal delay for a talkspurt to correspond to the difference in beginning times detected for that talkspurt.

24. The apparatus of claim 17, wherein the stereo decoding parameter expresses the estimated relative temporal delay between the first and second sound field signals in samples.

25. The apparatus of claim 17, wherein the stereo decoding parameter expresses an estimated angle of arrival based on the estimated relative temporal delay and the relative positioning of the first and second spatial locations.

26. The apparatus of claim 17, wherein the stereo decoding parameter corresponding to the encoded block of voice samples captured during a first time period is transmitted in the same packet as those voice samples.

27. The apparatus of claim 17, wherein the stereo decoding parameter corresponding to the encoded block of voice samples captured during a first time period is transmitted in a later packet than those voice samples.

28. The apparatus of claim 17, wherein the stereo decoding parameter corresponding to the encoded block of voice samples captured during a first time period is transmitted in a packet containing no encoded block of voice samples.

29. The apparatus of claim 17, wherein the stereo decoding parameter is transmitted once per talkspurt.

30. The apparatus of claim 17, wherein the method further comprises estimating, using voice samples captured in the approximate timeframe of the first time period, the signal energy in each sound field signal, and transmitting to the remote conferencing endpoint, in packet format, a stereo balance figure related to the relative signal energy in each sound field signal.

31. The apparatus of claim 17, wherein the method further comprises estimating, using voice samples captured in the approximate timeframe of the first time period, the signal energy in a frequency subband of each sound field signal, and transmitting to the remote conferencing endpoint, in packet format, a stereo balance figure related to the relative signal energy in that subband for each sound field signal.

32. A packet voice conferencing system comprising:

means for receiving concurrently-captured first and second sound field signals, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

means for encoding a digital data block to represent the combined first and second sound field signals captured within a first time period;

means for estimating, using the first and second sound field signals as captured in the approximate timeframe of the first time period, the relative temporal delay between the first and second sound field signals; and

means for encapsulating in a packet format the encoded digital data block and a stereo decoding parameter based on the estimated relative temporal delay.

33. The packet voice conferencing system of claim 32, wherein the means for receiving comprises a first sample buffer to receive digital voice samples representing the first sound field signal, and a second sample buffer to receive digital voice samples representing the second sound field signal.

34. The packet voice conferencing system of claim 32, wherein the means for receiving comprises a data link interface to receive digital voice samples from a remote conferencing endpoint.

35. The packet voice conferencing system of claim 32, wherein the means for encoding comprises:

an adder to create a combined sound field signal by summing the first and second sound field signals; and

an encoder to encode the combined sound field signal as created over an interval corresponding to the first time period, thereby encoding the digital data block;

36. The packet voice conferencing system of claim 32, wherein the means for estimating the relative temporal delay comprises a cross-correlator to correlate the first and second sound field signals for a plurality of relative time shifts.

37. A packet voice conferencing system comprising:

a sound field signal encoder to create a digitally-encoded signal block to represent both a first and a second sound field signal as captured within a first time period, the first and second sound field signals representing a single sound field captured at two spatially-separated points within a sound field;

a stereo parameter estimator to estimate the relative temporal delay between the first sound field signal and the second sound field signal within the approximate timeframe of the first time period; and

a packet formatter to encapsulate into at least one packet the digitally-encoded signal block and a stereo decoding parameter based on the estimated relative temporal delay.

38. The system of claim 37, further comprising a voice activity detector to detect when voice energy is represented in the first and second sound field signals, the voice activity detector supplying a voice activity detection signal to the packet formatter when voice activity is present, the packet formatter using the voice activity detection signal to inhibit packet generation when voice activity is not present.

39. The system of claim 38, the voice activity detector supplying the voice activity detection signal to the stereo parameter estimator, the stereo parameter estimator using the voice activity detection signal as an enabling signal.

40. The system of claim 38, the voice activity detector supplying the voice activity detection signal to the stereo parameter estimator as first and second signal components, the first component representing voice activity detection for the first sound field signal and the second component representing voice activity detection for the second sound field signal, the stereo parameter estimator estimating the relative temporal delay using the temporal delay between voice activity detection in the first and second components.

41. The system of claim 37, wherein the first and second sound field signals are digitally sampled, the system further comprising first and second sample buffers to respectively buffer digital samples for the first and second sound field signals and supply buffered samples to the stereo parameter estimator and sound field signal encoder.

42. The system of claim 37, wherein the sound field signal encoder comprises an adder to create a combined sound field signal by summing the first and second sound field signals; and

an encoder to encode the combined sound field signal as created over an interval corresponding to the first time period, thereby created the digitally-encoded signal block.

43. The system of claim 37, wherein the stereo parameter estimator comprises a cross-correlator to compute a first-to-second sound field signal cross-correlation coefficient for a plurality of relative time shifts, the estimated temporal delay based on the relative time shift

having the largest cross-correlation coefficient.

44. The system of claim 37, wherein the stereo decoding parameter comprises an arrival angle based on the estimated temporal delay and a known configuration of the two spatially-separated points.

45. The system of claim 37, wherein the stereo parameter estimator further comprises a signal energy estimator to estimate the signal energy present in each of the first and second sound field signals in the approximate timeframe of the first time period, the packet formatter encapsulating a stereo balance parameter related to the signal energy estimates.

46. The system of claim 37, wherein the stereo parameter estimator further comprises a signal energy estimator to estimate the signal energy present in a frequency subband of each of the first and second sound field signals in the approximate timeframe of the first time period, the packet formatter encapsulating a stereo balance parameter related to the signal energy estimates.

47. A packet voice conferencing system comprising:

a packet parser to receive voice packets received from a remote conferencing point, each voice packet containing at least one of an encoded signal block and a stereo decoding parameter;

a decoder to receive encoded signal blocks from the packet parser and decode those signal blocks to produce a voice sample stream; and

a playout splitter coupled to the voice sample stream, the splitter using the stereo

decoding parameter to create multiple output signal channels based on the voice sample stream.

48. The packet voice conferencing system of claim 47, further comprising a jitter buffer inserted in the voice sample stream between the decoder and the playout splitter.

49. The packet voice conferencing system of claim 47, wherein the stereo decoding parameter comprises a delay parameter, the splitter delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the delay parameter.

50. The packet voice conferencing system of claim 47, wherein the stereo decoding parameter comprises a balance parameter, the splitter modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the balance parameter.

51. The packet voice conferencing system of claim 50, wherein the playout amplitude modification is audio-frequency dependent.

52. The packet voice conferencing system of claim 47, further comprising a mixer to mix the output signal channels with other signal channels derived from voice packets received from another remote conferencing point.

53. The packet voice conferencing system of claim 52, further comprising a packet formatter to place the mixer output in packet format for transmission to a remote conferencing endpoint.

54. A packet voice conferencing system comprising:

means for decoding encoded signal blocks to produce a voice sample stream, each encoded signal block received in packet format from a remote conferencing point; and

means for splitting, based on the value of a stereo decoding parameter received in packet format from a remote conferencing point, the voice sample stream into multiple output signal channels to produce a stereophonic effect.

55. The packet voice conferencing system of claim 54, wherein the stereo decoding parameter comprises a delay parameter, the means for splitting the voice sample stream comprising means for delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the delay parameter.

56. The packet voice conferencing system of claim 54, wherein the stereo decoding parameter comprises a balance parameter, the means for splitting the voice sample stream comprising means for modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the balance parameter.

57. The packet voice conferencing system of claim 54, wherein the stereo decoding parameter

comprises an arrival angle parameter, the means for splitting the voice sample stream comprising means for calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the arrival angle parameter.

58. A packet voice conferencing method comprising:

receiving, from a remote conferencing point, a voice packet stream, at least some voice packets in the stream carrying a payload comprising an encoded signal block, at least some voice packets in the stream carrying a payload comprising a stereo decoding parameter;

decoding the encoded signal blocks to produce a voice sample stream;

splitting the voice sample stream into multiple output signal channels; and

manipulating the signal carried on at least one of the output signal channels based on the value of the stereo decoding parameter to create a stereophonic effect on the output signal channels.

59. The method of claim 58, wherein the stereo decoding parameter comprises a delay parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises delaying playout of the voice sample stream on at least one output signal channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the delay parameter.

60. The method of claim 58, wherein the stereo decoding parameter comprises a balance parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises modifying the playout amplitude of the voice sample stream on at least

one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the balance parameter.

61. The method of claim 58, wherein the stereo decoding parameter comprises an arrival angle parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the arrival angle parameter.

62. An apparatus comprising a computer-readable medium containing computer instructions that, when executed, cause a processor or multiple communicating processors to perform a method for packet voice conferencing, the method comprising::

receiving, from a remote conferencing point, a voice packet stream, at least some voice packets in the stream carrying a payload comprising an encoded signal block, at least some voice packets in the stream carrying a payload comprising a stereo decoding parameter;

decoding the encoded signal blocks to produce a voice sample stream;

splitting the voice sample stream into multiple output signal channels; and

manipulating the signal carried on at least one of the output signal channels based on the value of the stereo decoding parameter to create a stereophonic effect on the output signal channels.

63. The apparatus of claim 62, wherein the stereo decoding parameter comprises a delay parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises delaying playout of the voice sample stream on at least one output signal

channel, relative to playout of the voice sample stream on another output signal channel, based on the value of the delay parameter.

64. The apparatus of claim 62, wherein the stereo decoding parameter comprises a balance parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises modifying the playout amplitude of the voice sample stream on at least one output signal channel, relative to the playout amplitude of the voice sample stream on another output signal channel, based on the value of the balance parameter.

65. The apparatus of claim 62, wherein the stereo decoding parameter comprises an arrival angle parameter, and wherein manipulating the signal carried on at least one of the output signal channels comprises calculating a delay parameter for at least one output signal channel to create the perception that the audio signal represented in the voice sample stream is arriving at an angle corresponding to the arrival angle parameter.